

See by 1005
File: OPR

27 August 1974

TO: Director/OPR
RE: Critique of the Bayesian Analysis

Lew--

From where I sit, one of the principal advantages of the Bayesian approach is that it gives me confidence that each analyst is forced to deal with each piece of evidence in a systematic fashion and to make an explicit judgment about it. This advantage--though not peculiar to the Bayesian approach--does much for me not only as a person responsible for the analysis but also as a consumer of the judgments involved. The critique does not highlight this advantage.

I think the paper itself would benefit from a short summary of the principal findings--plus and minus. Perhaps a copy of the last substantive report could be attached to show the reader what we are discussing.

Finally, I feel the lack of some sort of discussion of where we go from here...A so-what, what's-next kind of thing. It does us little good to experiment with new techniques if we don't do something with them after we complete an experiment that seems to have gone well. I suggest adding a section describing the characteristics of the type of problem to which this technique is applicable and how we might identify such problems.


If you wish to discuss this further, I'm available.

signed
Ed Proctor

MEMORANDUM FOR: The DDI - A/DOE

Ed,

In view of your personal interest in the experiments applying Bayesian analysis to evaluate evidence on the likelihood of a major North Vietnamese attack in South Vietnam, I am forwarding the final draft of our critique of our experiment prior to publication. The critique is based upon responses from several dozen customers and participants, as well as upon the experiences of the OPR coordinators. Your comments will be welcome.


Lewis J. Lapham

23 August 1974

(DATE)

25X1

FORM NO. 101 REPLACES FORM 10-101
1 AUG 54 WHICH MAY BE USED.

(47)

C-O-N-F-I-D-E-N-T-I-A-L

Approved For Release 2005/12/24 : CIA-RDP80B01495R000700090011-7

OFFICE OF POLITICAL RESEARCH

22 August 1974

A CRITIQUE OF THE EXPERIMENTAL CIA AND USIB
BAYESIAN ANALYSES OF VIETNAM

1. The Office of Political Research initiated in late 1973 a Bayesian analysis on the likelihood of a major North Vietnamese attack on the South. This was an experimental project, using the services of selected experts in Vietnamese affairs in CIA. In February 1974, USIB directed OPR to conduct a parallel project on the same issue, but drawing upon analysts from those components of the Intelligence Community which wished to participate. Weekly progress reports were issued on each of the Bayes projects, from the beginning until 13 June.

2. The original aims of the Bayes project were: to test the validity and effectiveness of the Bayes formula as a tool for evaluating evidence on a major intelligence problem; to investigate the possible predictive or trend-indicating qualities of the Bayes method; to learn more about the willingness of qualitative analysts to make numerical judgments;

C-O-N-F-I-D-E-N-T-I-A-L

Approved For Release 2005/12/24 : CIA-RDP80B01495R000700090011-7

25X1

and to devise more effective presentational formats for intelligence reports.

3. The following critique discusses the achievements and failures of the project in respect to the stated goals, and touches on some problems which were not anticipated at the outset. There were also some unexpected or unsuspected advantages which emerged in the course of the experiment. On one important point, i.e., the predictive capability of Bayesian analysis, there was no dramatic showing one way or the other. Because no major military offensive was actually launched during the time of the exercise, there is no way of knowing whether the Bayes method would have revealed early trend warnings.

Positive Results of the Experiment

4. For most of the participants there were real gains in knowledge and understanding stemming from this exercise. Generally speaking, their initial attitudes ranged from skepticism to scarcely veiled hostility. Yet, in the course of the exercise analysts accustomed to qualitative expression of values acquired a genuine familiarity and confidence with the use of numbers to express substantive judgments. This is one of the more significant findings of the experiment, for it is

contrary to the oft-expressed assumption that "traditional" analysts are unable or unwilling to use quantitative indicators.

5. None of the analysts felt, in retrospect, that the mathematical revision of their prior estimates--based on their assigned likelihood ratios--differed significantly from their intuitive judgment of the situation. In the early stages several participants complained that the Bayes formula carried them down to levels ^{of probability (?)} where they felt uncomfortable. Their subjective judgments, in fact, tended to lag behind the Bayesian revision by one or two weeks. After this period of psychological adjustment, most analysts felt at ease at much lower positions on the graph. Thus, in this exercise at least, careful consideration of relevant data did revise expert estimates in the "right" direction. This, also, is a significant finding which tends to confirm the claim of Bayesian partisans that their method overcomes a natural conservative inclination of analysts to stick to a position even after the evidence dictates a change.

6. The procedural arrangements of the exercise were generally quite effective, both in terms of the individual analyst's facility in using Bayes as a means of expressing values, and as an aid in presentation of the results.

The exercise was conducted on the basis of individual contacts, rather than by meetings, and thereby avoided difficulties stemming from dominant personalities and pressures toward consensus. Indeed, no consensus was sought.

7. The process by which items of evidence were submitted, consolidated, and redistributed did result in a richer exchange of raw data than routinely takes place. Occasionally analysts found that they acquired an item via our consolidated list which they might not ordinarily have received, or they got it more quickly than it normally would have taken to reach their desk. In addition, different analysts normally place greater reliance on different types and sources of intelligence. The appearance of an item on the consolidated list provided the submitting analyst with a means of calling that item to the attention and careful consideration of his colleagues. Through this process, there was at least an implicit exchange of opinion among the various analysts over what evidence was relevant.

8. The format of the weekly progress reports was intended to be as ^{one or the other} clear, simple, and brief as possible. In this the experiment succeeded, judging from the responses of our principal consumers. The reader could take in at a glance the spread of opinion among senior

analysts on the intelligence question, and was shown a measure of central tendency to indicate the general trend in the collective opinion of the group. The weekly reports offered a sentence or two of Principal Trends, with no textual analysis. Items of evidence were submitted and evaluated by the experts, and listed in the report. The reader was left to observe how that data affected the weekly graph lines or to compare the results with his own evaluation of the evidence. On several occasions, our customers questioned how the various participants evaluated specific evidence to reach new positions on the graph. In general, those who paid particular attention to logistical and tactical intelligence tended to rate the chances of an offensive much higher than those who followed political intelligence.

9. Both the participants and the customers held varied views of the value and advisability of identifying the participants by name. The CIA project identified on the graph the name and office of each analyst, but because several of the USIB principals preferred anonymity, names were not used in the community version. Advocates of identification by name and office stressed the value of individual accountability, and held that more thorough and thoughtful assessments resulted from

this approach. The participants whose names appeared on the graph were commended by the DCI and others on their willingness to express openly their own judgments. Supporters of the opposing view stated that listing of names tended to personalize the exercise, i.e., readers paid more attention to the [] line," or the "[] line" than to the substantive issues of trends and evidence. The merits of this issue remain in dispute, but the majority of our readers seem to lean towards open identification of participants. 25X1

10. Of major practical import was the finding that participation in the Bayesian project did not, in general, consume great amounts of the working-level analysts' time. Most of the participants indicated that they spent a total of 1/2 to 2 hours weekly in the submission of evidence and assignment of likelihood ratios. There were exceptions on either end of this average time scale, which were generally reflective of the individual's interest in the project.

11. The procedural and substantive results of this weekly exercise gave rise to a number of spin-offs or side benefits. Some participants said they found

the weekly submission of evidence and the evaluation of all the items a useful review point within their own staff. Others used the trend of opinion in both the CIA and USIB exercises as a background for writing general analytical pieces about the Vietnam situation. Still other offices used the accumulated Progress Reports as a chronological archive of evidence when they had to write review articles or coordinate an NIE on the subject.

Potential Difficulties in Future Adaptations of Bayes

12. The Bayes method as applied in this subject would be difficult to adapt indiscriminately to indications analysis. If a situation was of sufficient importance to require expert evaluation on a daily-- or more frequent--basis, it is likely that the analysts most concerned would be heavily engaged in conventional reporting on the crisis, and would not be readily available for collecting and assessing data for a Bayesian project.

13. The main difficulty in applying Bayes to crisis situations probably lies in the administration of the technique. This might be overcome by designing an interactive computer program which would enable analysts to apply Bayesian and other probabilistic procedures for revising previous assessments on a current, "real-time"

basis. The development of such a program is, in fact, under consideration as a project for the near future in the Analytical Techniques Group of OPR.

14. The most serious problem affecting the validity of the Bayesian technique for revising estimates concerns the nature of the evidence collected by our Intelligence Community. Intelligence reporting follows a rule similar to journalistic reporting: adverse or unusual events (in this case, indications of hostilities) are prominently reported and non-events are not. In those weeks during this exercise when there was little relevant data or when the available data was not assessed as diagnostic, the analysts were "frozen" numerically at the level of their previous estimates when, in fact, that absence of data and/or the passage of time may have changed an analyst's judgment on the likelihood of an offensive in one direction or another. Our analysts, because of their experience with the data, satisfactorily resolved the problem by a sort of internal weighting process when they assigned likelihood ratios to the evidence. It would be desirable, however, in future Bayes projects to have some means of dealing with this problem in a more formal and rigorous way.

15. There was a further problem which occasionally came up in the submission of items of evidence for consideration: the same source of evidence was often described differently, or analysts would differ in their opinion of what parts of a source's content was most relevant (i.e., they would select certain facts from the body of a report and ignore other parts of it). This raised the question of whether a single report should be treated as a discrete item of evidence or whether its component parts could be dissected into several items. In addition, it was not always clear whether the analysts actually had the original document of evidence available for scrutiny, or whether they depended on our general description of an item as the basis for assigning their likelihood ratios.

16. There was also some question of the value of estimating the likelihood of an offensive (or any other event) against a fixed item deadline; in this case, the weather was probably a significant enough factor to justify a cut-off date (e.g., 30 June) beyond which a different set of conditions would affect the likelihood of the event. For the interests of the policymaker, however, it might have been more useful if we had established what is called a "floating window."

In this particular case, we could have asked the analysts each week to assess the evidence as it reflected the likelihood of a major offensive *in the next 90 days* (or the next 30 days). This would also deal directly with the problem that evidence is non-stationary and "decays" over time: that is, especially at the outset of a long project, any item of data is much more relevant to what will happen in the near future than to what may happen by some fixed date at great distance in the future.

17. There was an initial concern that, either because of our use of a graphic central tendency indicator or because individual estimates were identified by analysts' names, the participants would feel pressured to conform with the trend of the group's collective opinion. For a couple of important reasons, however, this did not become a problem in the Bayesian analysis project. First, the selection of experienced senior analysts, accustomed to defending their judgments on critical intelligence questions, assured a high degree of integrity in this project. A second, and more technical, point is that the participants each week were focusing their attention and quantitative evaluation on individual items of evidence, and not on the more

general question of the likelihood of an offensive. The avoidance of the problem of group pressures in this particular exercise, however, does not mean that it can be as easily dealt with in any future analyses of this kind.

A Fundamental Reminder

18. One of the essential requirements for any useful application of quantitative techniques and for most other methodologies is the availability of analysts with great substantive depth and expertise, both for the selection of relevant data and for the best numerical evaluation of those data. Probabilistic procedures for revising estimates on a major question such as this are not--and cannot be--a substitute for detailed, in-depth analysis of all the subtleties of the subject. Methodologies are good for dealing with aggregates of data or directional trends in analysis. Even so, without expert control over the inputs to data bases and the analytical design of research projects, methodology would become an end in itself rather than a tool in support of meaningful analysis.